Original Article

Trends in Hearing
Volume 23: 1–9
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2331216519887247
journals.sagepub.com/home/tia
SAGE

Comparing Three Established Methods for Tinnitus Pitch Matching With Respect to Reliability, Matching Duration, and Subjective Satisfaction

Patrick Neff[1,2], Berthold Langguth[1], Martin Schecklmann[1], Ronny Hannemann[3], and Winfried Schlee[1]

Abstract

The pitch of tinnitus sound is a key characteristic that is of importance to research and sound therapies relying on exact tinnitus pitch matches. The identification of this tinnitus pitch is a challenging task as there is no objective measurement available. During the tinnitus pitch-matching procedure, the participant identifies an external sound that is most similar to the subjective perception of the tinnitus. Several methods have been developed to perform this pitch-matching procedure with tinnitus sufferers. In this study, we aimed to compare the method of adjustment, the two-alternative forced-choice (2AFC) method, and the likeness rating (LR) with respect to reliability, matching duration, and subjective satisfaction. Fifty-nine participants with chronic tinnitus were recruited and performed five consecutive runs of tinnitus matching. The participants were randomized to the three different pitch-matching methods. The intraclass correlation coefficients were .67 for method of adjustment, .63 for 2AFC, and .69 for LR, which can be interpreted as good reliability for all the three methods. However, the 2AFC method revealed significant larger within-subject variability than the other measures. Across the five runs and the three different methods, all participants learned to perform the pitch matching faster and with better self-rated accuracy. Comparing the three pitch-matching methods, LR is more time consuming and the participants were less satisfied with the 2AFC method. Overall, the three pitch-matching methods show good reliability. However, we identified differential aspects for improvement in all methods, which are discussed in this article.

Keywords

tinnitus, tinnitus pitch matching, likeness rating, two-alternative forced choice, method of adjustment

Date received: 7 September 2018; revised 27 September 2019; accepted: 4 October 2019

Introduction

Tinnitus is the conscious perception of a sound in the absence of any physical source. It is estimated that about 5% to 15% of the population is chronically affected by tinnitus (Hoffman & Reed, 2004). This tinnitus sound is often described as a tone or noise with specific spectral characteristics that can be unique for each individual participant. Moreover, it has been shown that the perception of tinnitus can fluctuate in various situations and environments (Probst, Pryss, Langguth, & Schlee, 2016; Schlee et al., 2016). Currently, there is no objective measurement available that can determine the individual sound characteristics of the tinnitus. The assessment of the tinnitus sound characteristics has therefore to rely on the subjective description of the participant that matches the perceived tinnitus to an external sound as precisely as possible.

[1]Department of Psychiatry and Psychotherapy, University of Regensburg, Germany
[2]University Research Priority Program "Dynamics of Healthy Aging" University of Zurich, Switzerland
[3]Sivantos GmbH, Erlangen, Germany

Corresponding Author:
Winfried Schlee, Multidisciplinary Tinnitus Clinic, University of Regensburg, Universtätsstrasse 84, 93053 Regensburg, Germany.
Email: winfried.schlee@tinnitusresearch.org

A precise matching of the tinnitus sound is not only an important measure for the research toward a better understanding of the general and neuronal mechanisms underlying the tinnitus perception but also a key measure that enables well-adjusted individualized sound treatments to suppress or reduce tinnitus sound perception. In the recent years, research on sound therapies (for a review, see Searchfield, Durai, & Linford, 2017) and basic research on the temporary suppression of tinnitus (Fournier et al., 2018; Neff et al., 2017; Roberts, Moffat, Baumann, Ward, & Bosnyak, 2008) have steadily increased. Approaches relying on tinnitus pitch in both research branches are in need of a precise and reliable matching of the individual tinnitus pitch.

In parallel to these developments, studies on the methodology of tinnitus pitch matching have become steadily more numerous continuing to this day. Some of these studies merely applied standard audiometric methods (i.e., demonstration of hearing-relevant frequency sets) to assess tinnitus pitch—a practice which is still widespread in clinical routine and also generating data for concurrent studies (e.g., 31; Gollnast et al., 2017).

Besides and after this audiology-guided area of tinnitus pitch matching, the method of adjustment (MOA) emerged. In short, MOA methods allow for mostly user-controlled adjustment of the central parameters of tinnitus pitch and loudness. These parameters are controlled by mostly knob and slider and to a lesser degree button or graphical user interface (GUI) interaction. Instructions were mostly given beforehand or on-screen in the case of GUIs (e.g., Henry, Rheinsburg, & Ellingson, 2004b; Henry, Rheinsburg, Owens, & Ellingson, 2006; Tyler & Conrad-Armes, 1983).

Contrary to this user-guided method, other approaches have been developed where the loudness matching is taken care of algorithmically, usually precursing the tinnitus pitch matching of the predefined target frequencies. The two-alternative forced-choice method (2AFC; see Penner & Bilger, 1992) and the likeness rating (Norena, Micheyl, Chéry-Croze, & Collet, 2002; Roberts, Moffat, Baumann, Ward, & Bosnyak, 2006) approach are the most important examples for this algorithm-guided methodology. Following the principle of the 2AFC methodology, there are two sound examples presented to the participant who is then forced to pick one of the two examples, that is, more similar to the subjectively perceived tinnitus. After the participant has made the decision, a new pair of sound examples is played, and the participant has again to decide which example is more similar to the tinnitus. The sound examples are chosen in a way to narrow down the search interval to a frequency range that comes close to the individually perceived tinnitus sound with a small number of reversals. The algorithm for defining these sound examples underwent several modifications in the following years by different research groups. In the end, optimal step size for the central frequency domain emerged to be around 100 Hz (or 1/12 octave = 1 semitone when adjusted for the nonlinearity of physical frequencies behind musical scales or human auditory pitch perception; e.g., 56, Wunderlich et al., 2015).

As another algorithm-guided tinnitus pitch-matching methodology, two research groups independently introduced the method of rating standard audiometric frequencies for its contribution (i.e., likeness or similarity) to the perceived tinnitus (Norena et al., 2002; Roberts et al., 2006). The rating of likeness was performed not only on a 0–10 scale in subsequent studies (Basile, Fournier, Hutchins, & Hébert, 2013; Fournier & Hébert, 2012; Hébert & Fournier, 2017) but also on a percent scale (Hoare, Edmondson-Jones, Gander, & Hall, 2014; Roberts et al., 2008). Beyond that, the rating on the percent scale almost exclusively was performed in decades (e.g., 10%, 20%, or 90%). In the study of Norena et al. (2002), LR and the absolute hearing thresholds were overlayed and the authors observed a relationship between the shapes of both curves in that regions with the most pronounced hearing loss coincide with elevated level of tinnitus pitch likeness. The advantage of the LR can be seen in its ability to depict the tinnitus pitch likeness over the whole relevant frequency spectrum, thus giving an array of probabilities, instead of narrowing down the tinnitus pitch to a single frequency as performed in MOA or 2AFC. Still, while the LR method implicates that there may be no relation between the tinnitus pitch likeness, the method could be used to narrow down the search space to identify the tinnitus pitch (Norena et al., 2002) or even extract pitch matches from the LR results (Hébert, 2018).

Evaluating the LR method, Hoare et al. (2014) repeated the procedure at different time intervals resulting in an acceptable test–retest reliability with a 2 week but not with a 3-month interval. A further study was directly comparing the LR with the 2AFC method (Hébert, 2018) with the specific aim to extract one dominant pitch and loudness matching for 2AFC and accordingly three dominant matches for LR. The matching was repeated two times at a 1-month interval. Results were indicative of a superior test–retest reliability of LR compared with 2AFC.

The aim of this study was to compare these three established methods, namely, MOA, 2AFC, and LR. For this comparison, three evaluation categories have been of particular focus: reliability, matching duration, and satisfaction. The reliability of the tinnitus pitch matching is important for basic research as well as for clinical treatments with sound therapies. The duration of the tinnitus matching is of practical importance for the clinical routine. Since the matching of the tinnitus pitch is purely subjective, the self-rated satisfaction with the matching result is an important feedback of the tinnitus individual that can be used as an additional indicator for the precision of the matching.

Methods

Participants

We recruited a sample of 59 tinnitus participants from the interdisciplinary tinnitus clinic at the university hospital Regensburg with an age range spanning from 18 to 75 years. Convenience sampling was applied with the following inclusion and exclusion criteria: Primary and sole inclusion criterion was chronic, tonal tinnitus (single pitch) present for at least 6 months. Exclusion criteria were neurological or psychiatric diseases, concurrent tinnitus interventions, substance abuse, hearing aids, and finally hearing loss above 40 dB at any frequency up to 8 kHz. All participants gave written informed consent after being informed about the scope and procedural details of the study. The study was approved by the ethical review board of the University of Regensburg (approval number 17-658-101). Demographic characteristics of the participant sample are described in Table 1.

To ensure comparability between the pitch-matching methods, participants were randomly assigned to three groups with the goal of three equivalent groups matched for age, sex, hearing loss, and musicality. The resulting groups did not show any statistically significant difference with respect to age (t test, $p > .1$), sex (χ^2 test, $p > .9$), hearing loss (t test, $p > .4$), or musicality (χ^2 test, $p > .8$). Beyond these primary matching parameters, we also report nonsignificant differences in further assessed variables relevant to the study procedure (t tests), namely, educational status ($p > .8$), tinnitus duration ($p > .6$), self-reported subjective tinnitus loudness ($p = .19$), and time aware of tinnitus ($p > .4$).

Questionnaires

Upon the actual experiment, participants filled in an online questionnaire comprising the Tinnitus Sample Case History Questionnaire for clinical and demographic data (Langguth et al., 2007), a short version of the Tinnitus Questionnaire (mini-TF, Goebel & Hiller, 1994), and the German adaption of the Tinnitus Handicap Inventory (Newman, Jacobson, & Spitzer, 1996). Questions, comments, and ratings during the experimental procedure were assessed with paper and pencil.

Table 1. Demographic Characteristics of the Study Population (Newman et al., 1996).

Demographic characteristics	Study population ($N = 59$)
Age (years)—mean ± *SD*	53.9 ± 9.0
Sex—*n* (%)	
Male	38 (64.4)
Female	21 (35.6)
Average hearing loss (dB)—mean ± *SD*	18.4 ± 18.8
THI sumscore (0–100)—mean ± *SD*	55.7 ± 11.4
Musical experience—*n* (%)	
No musical experience	44 (74.6)
HM	15 (25.4)
Musical practice hours per week (HM, hours)—mean ± *SD*	0.68 ± 0.48

Note. HM = hobby musician; *SD* = standard deviation; THI = Tinnitus Handicap Inventory.

Audiometry

Hearing thresholds were measured in the frequency range from 125 Hz to 8 kHz in octave steps with semi-octave steps between 0.5 and 1 (i.e., 0.75 kHz), 1 and 2 (i.e., 1.5 kHz), 2 and 4 (i.e., 3 kHz), and 4 and 8 kHz (i.e., 6 kHz), respectively (Madsen Midimate 622 D; GN Otometrics, Denmark) with Sennheiser HDA 2000 headphones (Sennheiser, Germany).

Study Design

At the beginning of the experiment, the participants were informed about the study procedures and signed the informed consent. All participants performed five consecutive runs of tinnitus matching. Between the runs, participants had a break of at least 5 min where they could read news, solve crosswords, or sudoku for distraction. After the fifth session of the experiment, participants filled in the online survey with the questionnaires described in the earlier questionnaire section. Upon completion of the survey, participants underwent pure tone audiometry (Figure 1). Finally, participants were debriefed and dismissed. The experiment lasted about 90 min on average. No measurements had to be excluded.

Tinnitus Pitch-Matching Methods

As a first measure, as common in tinnitus pitch matching, an ear was defined on which the matching sounds were presented (Henry & Meikle, 2000). Ideally, the ear contralateral to the tinnitus was chosen in the case of unilateral tinnitus and good hearing in the contralateral ear. In case of bilateral tinnitus with no preference to one side, the matching was performed on the better hearing ear (the ear with less average hearing loss over all tested

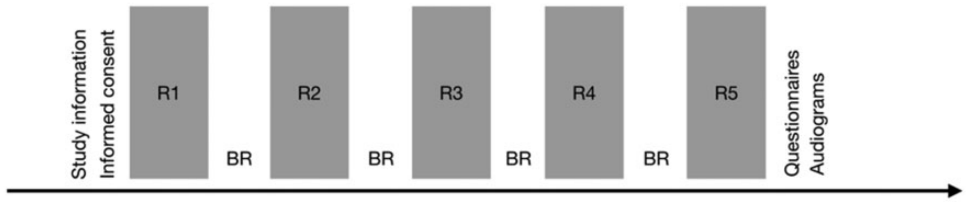

Figure 1. Flowchart of the experimental procedure. There were five consecutive runs (R1–R5) for tinnitus matching, interrupted by 5-min BR to relax and distract the participants. BR = break.

frequencies). Finally, if all of the above options failed, participants were able to choose their preferred ear for matching. Upon decision on the matching ear, all of the three methods were configured to present their sounds on the respective ear exclusively. With respect to presentation sound levels (i.e., loudness), LR was user-driven as the loudness could be adjusted for each frequency in each trial with a slider starting from a just audible level. In MOA and 2AFC, levels were adjusted to a comfortable level (see details in the following subsections).

Method of adjustment. After a 500 Hz tone had been adjusted to a comfortable frequency, participants were instructed to use a rotary encoder to adjust the frequency of the matching pure tone to the pitch of their tinnitus. It was emphasized and demonstrated that the rotary encoder can be used for both fast scrolling through the whole audible spectrum as well as slowly turned for fine tuning. Following this central step of pitch matching, octave confusion was tested with a respective switch. If an octave confusion was identified, participants were asked to redo the pitch-matching procedure. Finally, after successful pitch matching, the loudness of the matching sound was adjusted to match the loudness of the tinnitus.

Two-alternative forced choice. The 2AFC procedure was done in three steps: First, a coarse definition of the octave where tinnitus is most probably situated was defined. This was achieved by both having an eye on the audiometric profile and testing the limits of the range with probe tones. The latter was performed in our case and is comparable to the method of limits (Tyler & Conrad-Armes, 1983). The upper and lower limits of this octave then served as the extreme of the starting bracket of the double stair case (e.g., 4000 and 8000 Hz, respectively). This bracket then served for the actual tinnitus pitch matching, where the final frequency was approached on the double staircase in one-third octave steps with a maximum of seven iterations per run. Finally, as a third step, octave confusion was tested and procedure repeated, in case of actual confusion. This last step was comparable to the procedure in MOA.

Likeness rating. A frequency list of 11 frequencies (0.5, 1, 2, 3, 4, 5, 6, 7, 8 ,10, and12 kHz according to Hoare et al., 2014) was displayed on the operators GUI and were presented in sequence from top to bottom. Upon button press of a frequency, the sound was played for 3 s. First, participants were instructed to adjust the level of the sound to the loudness of their tinnitus. Following that, participants rated the likeness of the presented sound to the subjective tinnitus on a percent scale. Upon completion, the next frequency was presented and the procedure continued until all 11 frequencies were adjusted and rated. The procedure was implemented in Matlab as a GUI application controllable via computer mouse by the operator (study personnel) and via volume fader by the participants. The set of the 11 frequencies spanning up to 12,000 Hz was pseudorandomly generated so that no direct neighbor frequency was presented in sequence, and that the single runs did not start or end with identical frequency to counteract anchor and other learning effects. For all of the three methods, a final best matching frequency was chosen. Therefore, for LR, participants had to opt for a favorite frequency if the same LR was given for several frequencies.

Statistical Analysis

Statistical analysis was performed with R version 3.3.3 (R project, Vienna, Austria). Several measures were used to assess the reliability of the three tinnitus pitch-matching methods. The intraclass correlation (ICC) was calculated using the "irr" library (version 0.84). The coefficient of variation (CV, also known as relative standard deviation) was calculated as a ratio between the standard deviation and the mean. Furthermore, the CV compliance rate (CVCR) was calculated to identify the percentage of participants with a CV below a given cutoff value. For this cutoff value, we chose the criterion <0.33, as this is commonly interpreted as an acceptable CV (Ruhe, Fejer, & Walker, 2010). All these measures are based on a linear frequency scale, which does not respect the natural pitch perception of the human ear, which is better described as proportional to the logarithm of the frequency. Therefore, we also calculated the difference between the matched tinnitus pitches in octaves. For each, an average difference between all matching results—measured in octaves—was calculated. Mixed-model analyses of variance (ANOVAs) were calculated with the "nlme" library (version 3.1-131) modeling a random intercept per participant.

Results

Tinnitus Pitch-Matching Results

Figure 2 shows the pitch-matching results for each participant and run. The CV was calculated across the five measurements for each participant, and an average CV is reported for each method in Table 2. The highest CV was found for the 2AFC method with 43.6%, which was significantly higher than for the LR (two-sample t test, $t = 2.20$, $p = .038$). There was no significant difference between the CV values of the MOA and the 2AFC method ($p > .1$) nor between the MOA and the LR ($p > .5$). In addition, the CVCR compliance rate (CVCR) was calculated for each method with a cutoff criterion of CV <.33 (Ruhe et al., 2010). The largest CVCR was found for the MOA with 80% of the participants showing a CV below .33. Of the participants using LR, 73.7% scored below this level, while only 55% of the participants using the 2AFC method reached such a low CV value. For each participant, the mean difference between the five pitch-matching results was calculated in octaves. The average values and standard differences for each method are reported in Table 2. The largest average was found for the 2AFC method with a mean difference of 1.07 octaves. This average was significantly higher than the average of the MOA (two-sample t test, $t = 2.31$, $p = .03$) and significantly higher than the LR (two-sample t test, $t = 2.37$, $p = .03$). There was no significant difference in the mean octave differences between the MOA and the LR ($p > .9$). The average tinnitus pitch measured with the MOA was with 4,697 Hz significantly lower than the average pitch measured with the 2AFC method (two-sample t test, $t = 2.90$, $p = .007$) and also significantly lower than the average pitch measured with the LR (two-sample t test, $t = 3.46$, $p = .001$). There was no significant difference in the average pitch measures between the 2AFC method and the LR ($p > .8$).

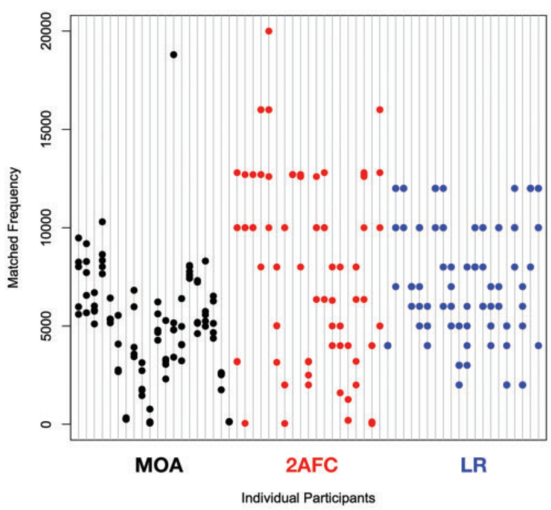

Figure 2. Dot plot of the tinnitus matching results for the 59 tinnitus participants. The dots represent individual pitch-matching results. Each individual participant performed five consecutive measures of the tinnitus pitch, visualized by the five dots on the respective line. MOA measurements are shown in black, the 2AFC measurements in red, and the LR measurements in blue. MOA = method of adjustment; 2AFC = two-alternative forced choice; LR = likeness rating.

Duration of Pitch Matching

The time duration for the performing the pitch matching was measured for each run and each tinnitus participant, and the mean durations are shown in Table 3. A mixed-model ANOVA revealed a significant main effect for the run ($F = 144.4$, $p < .0001$) and the method type ($F = 12.5$, $p < .0001$), while the interaction effect of run and method was not significant ($p > .1$). Across all pitch-matching methods, the participants learned quickly to perform the pitch matching with shorter time durations. For all methods, the comparison between Run 1 and Run 5 shows much faster pitch matching for the last session (paired t tests, all $p < .0001$). Post hoc analysis on the main effect for the method type showed that the LR was always the method with the longest duration. In all the five runs, the LR was the significantly slower than the fastest method (all $p < .01$). Between the 2AFC

Table 2. Reliability Measures and Average Pitch Measures for the Three Different Pitch-Matching Methods.

Method	ICC (95% CI)	Average CV (%)	CVCR (%)	Mean OD	Mean frequency (Hz)
MOA	0.67 [0.50, 0.83]	28.4	80	0.42 ± 0.36	4697
2AFC	0.63 [0.44, 0.80]	43.6	55	1.07 ± 1.20	7779
LR	0.69 [0.51, 0.84]	23.4	73.7	0.41 ± 1.20	7632

Note. ICC = intraclass correlation; CI = confidence interval; CV = coefficient of variation; CVCR = CV compliance rate; OD = octave difference; MOA = method of adjustment; 2AFC = two-alternative forced-choice; LR = likeness rating.

Table 3. Time Duration for Pitch Matching, Measured in Seconds (mean ± *SD*).

Method	Run 1	Run 2	Run 3	Run 4	Run 5
MOA	385 ± 205	164 ±73	146 ± 111	152 ± 81	131 ± 81
2AFC	327 ± 108	215 ± 89	205 ± 86	170 ± 60	161 ± 78
LR	480 ± 162	335 ± 132	256 ± 97	252 ± 77	233 ± 72

Note. MOA = method of adjustment; 2AFC = two-alternative forced choice; LR = likeness rating.

Table 4. Subjective Self-ratings of the Participants on the Accuracy of the Pitch Matching (mean ± *SD*).

Method	Run 1	Run 2	Run 3	Run 4	Run 5
MOA	8.1 ± 0.78	8.2 ± 1.24	8.45 ± 1.23	8.45 ± 1.05	8.7 ± 1.03
2AFC	7.5 ± 1.43	6.6 ± 1.9	7.9 ± 1.12	7.35 ± 1.81	7.9 ± 1.37
LR	8.21 ± 1.23	8.79 ± 0.79	8.84 ± 1.07	8.68 ± 0.89	8.63 ± 1.11

Note. Range: 1–10, 1 = *not satisfactory at all*, 10 = *highly satisfactory.* MOA = method of adjustment; 2AFC = two-alternative forced choice; LR = likeness rating; *SD* = standard deviation.

method and MOA, there was no significant difference found in neither of the runs (all $p > .05$).

Subjective Satisfaction With Matching Accuracy

After each pitch matching, the participants were asked to rate the matching accuracy on a scale from 0 to 10. The mean values and standard deviation for all methods and runs are given in Table 4. A mixed-model ANOVA on these self-rating values revealed a main effect for the run ($F = 9.6$, $p = .002$) and the method type ($F = 9.1$, $p < .001$), but no significant interaction ($p > .6$). The main effect for the run reflects the tendency that the satisfaction slightly increased over the five consecutive runs. However, post hoc analysis between the first and the fifth run revealed only for the MOA a significant improvement (t test, $t = 2.07$, $p = .046$). The main effect for the method reveals that participants using the 2AFC method always gave the lowest ratings across all runs, while the participants using the LR gave the highest ratings in four of the five runs. In Runs 2, 3, and 4, the difference between the satisfaction self-ratings of the 2AFC method and the LR reached statistical significance (all $p < .011$).

Discussion

MOA, 2AFC, and LR are three different pitch-matching methods that have been compared on 59 chronic tinnitus participants. The pitch-matching methods have been compared with respect to their retest reliability, the time duration for performing the matching procedure, and the subjective satisfaction of the participants with the matching result.

To evaluate the reliability of pitch-matching methods, we used four different measures highlighting different aspects of the retest results. The ICC was calculated as a commonly used measure for retest reliability with multiple repeated measures. The ICC values of all the three pitch-matching methods (Table 2) can be interpreted as good reliability. End points of the confidence intervals extended between fair and excellent ICC values. Since the ICC measures resulted in wide confidence intervals, it was not possible to decide whether there is one method significantly less or more reliable than the others. Similar observations for reliability were made in several former studies where different matching methods were compared (Basile et al., 2013; Hauptmann et al., 2016; Henry, Flick, Gilbert, Ellingson, & Fausti, 2004a; Tyler & Conrad-Armes, 2009; Wunderlich et al., 2015). Conflicting results were shown in an other study demonstrating superior test–retest concordance of LR in comparison to 2AFC (Hébert, 2018). To test for differences not accessible with the ICC method, we also calculated the CV, which is a measure for the relative standard deviation of the matching results. The CVs for participants using the 2AFC method were found to be much higher than in the participants using the LR. This also reflected in the CVCR. Only 55% of the participants using the 2AFC method were able to produce a CV smaller than .33. On the other side, the CVCR for the LR reached 73.7% and the MOA 80%. A similar analysis was performed in Hauptmann et al. (2016) where the comparison of 2AFC and MOA in matching a test tone resulted in 80% of the trials within a 5% pitch interval for 2AFC and only 40% for MOA. Notably, given the specific task of matching to a fixed external sound, the results cannot be directly compared with the results in this study and are furthermore in conflict with the view of none or only minor differences in reliability between the established methods. Another important measure, for example, for the individual adjustment of sound therapies, is the mean octave difference. As some contemporary sound therapies (e.g., the notched music therapy or the notched hearing aid) need to adjust a notch (e.g., 0.5–1 octave) around the individually measured tinnitus pitch, the average deviation is of practical importance for the clinical treatment. We found that the mean octave difference in participants using the 2AFC method is significantly higher than in participants using the MOA or the LR. This could be explained by a rather low number of reversals applied in our study (i.e., 7) or a latent, systematic bias in the 2AFC experimental group.

The evaluation of the duration for tinnitus matching revealed a strong learning effect of all participants across all pitch-matching methods. The average matching duration in Run 5 was always more than 50% faster than in Run 1. This very fast learning effect will need to be considered for designing future studies with repeated tinnitus pitch matchings. In addition, we found that the LR method consistently needed a longer time duration for pitch matching, which can be explained by the time cost of its inherent procedurality (i.e., loudness matching of all probe tones). This dependence on procedural details as well as no previous studies testing multiple runs of different matching methods in parallel obstructs a meaningful discussion of testing duration. Fittingly, Henry et al. (2004a, page 134) noted for LR that

> However, note that for the Subject-Guided method, the time required to obtain thresholds and loudness matches at each frequency was not factored into the time of testing. Thus obtaining a pitch match with this method would take much longer if total testing time was combined.

Taken together, the duration of tinnitus matching lacks a proper conceptualization as details inherent to the procedure or dependencies between matching procedures or audiometric measures limit the measurement of *actual* matching duration.

To assess the subjective satisfaction of the participant with the pitch-matching result, the participants were asked to self-rate on a scale between 0 and 10 how much the matched tinnitus tone corresponds to the subjectively perceived tinnitus. We found that subjective satisfaction slightly increased from the first to the fifth run across all pitch-matching methods, especially MOA. This can be interpreted as an indicator that the participants not only learned to perform the pitch matching faster but also learned to match their tinnitus with better accuracy. A comparable slight increase overtime or certainly between the first and the subsequent session was found for MOA (Henry et al., 2004b) and for LR or 2AFC (Hébert, 2018). However, the improvement is rather small in magnitude in our data as well as in former studies. More studies will be needed to examine the learning progress in more detail. Furthermore, the analysis revealed that the participants in the 2AFC group were on average less satisfied with matching results than the participants using the MOA or the LR. This prima facie contradicts our findings, but could again be explained by the latent limitations of the 2AFC method in our study. Future studies could also profit from a differential set of questions regarding satisfaction such as comprehensibility, ease of use, certainty about the result, and comfort level of the procedures (Wunderlich et al., 2015).

In our study, we presented the matching sound stimulus to the ear contralateral to the tinnitus ear according to common practice. Yet, there is conflicting evidence and recommendations. Tyler and Conrad-Armes (1983) observed lower pitch matchings in some subjects in the contralateral matching procedure and recommend the use of ispilateral stimulation to "avoid any effects of diplacusis." Furthermore, this study identified seven to nine runs as the optimal number of repetitions in tinnitus pitch matching and proposed to track the variability of the results. Related to that, consistent measure over larger time intervals are needed to both identify persons with fluctuating tinnitus but also better prepare any study or treatment dependent on tinnitus pitch (Tyler, 1985). With increasing repetitions and related exposure to sounds, the effect of those sounds on the tinnitus itself but also the matching procedure have to be considered (Henry & Meikle, 2000; Tyler, 2005). Unfortunately, we cannot provide the reader with any data on such effects at this point.

Summarizing the results of this comparison, there was no pitch-matching method that is the clear winner in all categories. Future methodological studies on tinnitus pitch matching may take advantage of these results by developing a combined method melting the advantages of each method together (e.g., as proposed in Hauptmann et al., 2016). MOA was found to be a method with good reliability of the tinnitus matching, low variability of matching results, short time duration for the tinnitus matching, and high participant satisfaction. However, it has to be highlighted here that the frequencies of the matched tinnitus tones were significantly lower than the frequencies that have been matched with 2AFC or LR. With implementation that we used in this study, the participants using the MOA always started with an initial setting of 500 Hz and were asked to increase the frequency until they reach their individual tinnitus pitch. We suspect that this initially low frequency setting biased the participants toward lower frequencies and offering an anchor for their matching decision, which might have favored octave confusions toward lower frequencies. Future developments will need to address this disadvantage by better solutions for the initial setting (e.g., starting with a random frequency). The 2AFC method was found to be a fast technique for tinnitus pitch matching. The analysis of the ICC demonstrates a good reliability. The analysis of the CV as well as the mean octave difference, however, reveals large variations of the matched frequency within repeated measures of the individuals. This is most likely due to the algorithm that forces the participants into making a series of comparative decisions. If the participant makes a wrong decision in the beginning, this leads the following decision tree into a wrong direction (Wunderlich et al., 2015). Accordingly, the subjective satisfaction of the participants remained poor compared with the other matching procedures. Pitch matching with the LR was done with good reliability and relatively low variation of pitch-matching results. The subjective satisfaction of the participants with the results was relatively high. However, the participants needed significantly more time to perform the pitch matching compared with the other methods. An important limitation of the LR is the frequency resolution of the results. In our implementation, we used 11 different frequencies for the LR. The maximum and minimum frequencies as well as the frequency resolution are dependent on these predefined frequencies. In addition, more test frequencies would prolong the matching procedure. This is an immanent trade-off of the methodology. The researcher or clinician performing the LR therefore has to decide on the needed frequency resolution and time commitment.

Conclusion

Altogether, the compared methods for pitch matching show good reliability with acceptable matching durations and participant satisfaction. However, in all the aforementioned methods, we identified room for improvement. Beyond that, the meaningful combination of the three methods could improve reliability, matching duration, and satisfaction with the results. Especially in a time of emerging auditory treatments that depend on precise tinnitus pitch matching, future advancements are needed to develop methods that can be performed fast and with high reliability. This will help to improve the efficacy of the clinical treatment and also enable new insights in the scientific understanding of tinnitus.

Acknowledgments

The authors want to thank Lisa Zielonka, Susanne Staudinger, and Johannes Arnds for their extremely valuable support in the collection of the data.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Patrick Neff https://orcid.org/0000-0003-3174-4910

References

Basile, C. É., Fournier, P., Hutchins, S., & Hébert, S. (2013). Psychoacoustic assessment to improve tinnitus diagnosis. *PLoS One*, *8*(12), e82995. doi:10.1371/journal.pone.0082995

Fournier, P., Cuvillier, A. F., Gallego, S., Paolino, F., Paolino, M., Quemar. A., . . . Norena, A. (2018). A new method for assessing masking and residual inhibition of tinnitus. *Trends in Hearing*, *22*, 2331216518769996. doi:10.1177/2331216518769996

Fournier, P., & Hébert, S. (2012). Gap detection deficits in humans with tinnitus as assessed with the acoustic startle paradigm: Does tinnitus fill in the gap? *Human Auditory NeuroImaging*, 1–8. doi:10.1016/j.heares.2012.05.011

Goebel, G., & Hiller, W. (1994). The tinnitus questionnaire. A standard instrument for grading the degree of tinnitus. Results of a multicenter study with the tinnitus questionnaire. *HNO*, *42*(3), 166–172.

Gollnast, D., Tziridis, K., Krauss, P., Schilling, A., Hoppe, U., & Schulze, H. (2017). Analysis of audiometric differences of patients with and without tinnitus in a large clinical database. *Frontiers in Neurology*, *8*(Suppl.), 31. doi:10.3389/fneur.2017.00031

Hauptmann, C., Wegener, A., Poppe, H., Williams, M., Popelka, G., & Tass, P. A. (2016). Validation of a mobile device for acoustic coordinated reset neuromodulation tinnitus therapy. *Journal of the American Academy of Audiology*, *27*(9), 720–731. doi:10.3766/jaaa.15082

Hébert, S. (2018). Individual reliability of the standard clinical method vs patient-centered tinnitus likeness rating for assessment of tinnitus pitch and loudness matching. *JAMA Otolaryngology–Head & Neck Surgery*, *144*(12), 1136–1144. doi:10.1001/jamaoto.2018.2416

Hébert, S., & Fournier, P. (2017). Clinical validation of a new tinnitus assessment technology. *Frontiers in Neurology*, *8*(8), 711–718. doi:10.3389/fneur.2017.00038

Henry, J. A., Flick, C. L., Gilbert, A., Ellingson, R. M., & Fausti, S. A. (2004a). Comparison of manual and computer-automated procedures for tinnitus pitch-matching. *Journal of Rehabilitation Research and Development*, *41*(2), 121–138. doi:10.1682/jrrd.2004.02.0121

Henry, J. A., & Meikle, M. B. (2000). Psychoacoustic measures of tinnitus. *Journal of the American Academy of Audiology*, *11*(3), 138–155.

Henry, J. A., Rheinsburg, B., & Ellingson, R. M. (2004b). Computer-automated tinnitus assessment using patient control. *The Journal of Rehabilitation Research and Development*, *41*(6), 871–818. doi:10.1682/JRRD.2003.10.0158

Henry, J. A., Rheinsburg, B., Owens, K. K., & Ellingson, R. M. (2006). New instrumentation for automated tinnitus psychoacoustic assessment. *Acta Oto-laryngologica Supplementum*, *126*(556), 34–38. doi:10.1080/03655230600895291

Hoare, D. J., Edmondson-Jones, M., Gander, P. E., & Hall, D. A. (2014). Agreement and reliability of tinnitus loudness matching and pitch likeness rating. *PLoS One*, *9*(12), e114553–18. doi:10.1371/journal.pone.0114553

Hoffman, H. J., & Reed, G. W. (2004). Epidemiology of tinnitus. *Tinnitus: Theory and Management*. Retrieved from http://books.google.com/books?hl=en&lr=&id=BqEq9Re3L5UC&oi=fnd&pg=PA16&dq=Epidemiology+of+tinnitus&ots=ekeok_9xdc&sig=FcG1RiDRLuY4Gpfj8_24H4CTSsA

Langguth, B., Goodey, R., Azevedo, A., Bjorne, A., Cacace, A., Crocetti, A., ... Vergara, R. (2007). Consensus for tinnitus patient assessment and treatment outcome measurement: Tinnitus Research Initiative meeting, Regensburg, July 2006. *Progress in Brain Research 166*, 525–536. doi:10.1016/S0079-6123(07)66050-6

Neff, P., Michels, J., Meyer, M., Schecklmann, M., Langguth, B., & Schlee, W. (2017). 10 Hz amplitude modulated sounds induce short-term tinnitus suppression. *Frontiers in Aging Neuroscience*, *9*, 215–211. doi:10.3389/fnagi.2017.00130

Newman, C. W., Jacobson, G. P. & Spitzer, J. B. (1996). Development of the tinnitus handicap inventory. *Archives of Otolaryngology–Head & Neck Surgery*, *122*(2), 143–148. doi:10.1001/archotol.1996.01890140029007

Norena, A., Micheyl, C., Chéry-Croze, S., & Collet, L. (2002). Psychoacoustic characterization of the tinnitus spectrum: Implications for the underlying mechanisms of tinnitus. *Audiology & Neuro-otology*, *7*(6), 358–369. doi:10.1159/000066156. Retrieved from http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=12401967&retmode=ref&cmd=prlinks

Penner, M. J., & Bilger, R. C. (1992). Consistent within-session measures of tinnitus. *Journal of Speech and Hearing Research*, *35*(3), 694–700. doi:10.1044/jshr.3503.694. Retrieved from http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=1608262&retmode=ref&cmd=prlinks

Probst, T., Pryss, R., Langguth, B., & Schlee, W. (2016). Emotion dynamics and tinnitus: Daily life data from the "TrackYourTinnitus" application. *Nature Publishing Group*, 1–9. doi:10.1038/srep31166

Roberts, L. E., Moffat, G., Baumann, M., Ward, L. M., & Bosnyak, D. J. (2008). Residual inhibition functions overlap tinnitus spectra and the region of auditory threshold shift. *Journal of the Association for Research in Otolaryngology*, *9*(4), 417–435. doi:10.1007/s10162-008-0136-9

Roberts, L. E., Moffat, G., & Bosnyak, D. J. (2006). Residual inhibition functions in relation to tinnitus spectra and auditory threshold shift. *Acta Oto-laryngologica Supplementum*, *126*(556), 27–33. doi:10.1080/03655230600895358

Ruhe, A., Fejer, R., & Walker, B. (2010). The test-retest reliability of centre of pressure measures in bipedal static task conditions—A systematic review of the literature. *Gait & Posture*, *32*(4), 436–445. doi:10.1016/j.gaitpost.2010.09.012

Schlee, W., Pryss, R. C., Probst, T., Schobel, J., Bachmeier, A., Reichert, M., ... Langguth, B. (2016). Measuring the moment-to-moment variability of tinnitus: The track your tinnitus smart phone app. *Frontiers in Aging Neuroscience*, *8*, 294. doi:10.3389/fnagi.2016.00294

Searchfield, G. D., Durai, M. & Linford, T. (2017). A state-of-the-art review: Personalization of tinnitus sound therapy. *Frontiers in Psychology*, *8*, 1599. doi:10.3389/fpsyg.2017.01599

Tyler, R. S. (1985). Psychoacoustical measurement of tinnitus for treatment evaluations. *New Dimensions in Otorhinolaryngology Head and Neck Surgery*, 455–458.

Tyler, R. S. (2005). The psychoacoustical measurement of tinnitus. *Tinnitus Handbook*, 174–179.

Tyler, R. S., & Conrad-Armes, D. (1983). Tinnitus pitch: A comparison of three measurement methods. *British Journal of Audiology*, *17*(2), 101–107. Retrieved from http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=6626779&retmode=ref&cmd=prlinks

Tyler, R. S., & Conrad-Armes, D. (2009). Tinnitus pitch: A comparison of three measurement methods. *British Journal of Audiology*, *17*(2), 101–107. doi:10.3109/03005368309078916

Wunderlich, R., Stein, A., Engell, A., Lau, P., Waasem, L., Shaykevich, A., ... Pantev, C. (2015). Evaluation of iPod-based automated tinnitus pitch matching. *Journal of the American Academy of Audiology*, *26*(2), 205–212. doi:10.3766/jaaa.26.2.9